Reavis, W. C. "The importance of study programs for high-school pupils," School Review, 19:398-405, June, 1911.
Robert, A. C. "Supervised study in the Everett High School," School Review, 24:735-45, December, 1916.
Rynearson, Edward. "The conference hour in the Pittsburgh high schools," School Review, 20:246-53, April, 1912.
Stetson, Paul C. "A survey of supervised study," The American School Board Journal, 54:19-20, 85-86, June, 1917.
Wiener, William. "Home-study reform," School Review, 20:526-31, October, 1912.
Wiener, William. "The social value of school study versus home study," Chapter XI, The Modern High School, edited by Chas. H. Johnston. New York: Charles Scribner's Sons, 1916, p. 295-311.
Willett, G. W. "Supervised study in high school," School Review, 26:259-72, April, 1918.
Young, Eula, and Simpson, M. R. "A technique for the lengthened period," School Review, 30:199-204, March, 1922.

BULLETIN NO. 27

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

EFFECT OF PRACTICE ON INTELLIGENCE TESTS

By

H. N. Glick
Professor of Educational Psychology,
Massachusetts Agricultural College
Amherst, Massachusetts

PRICE 30 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1925


UNIVERSITY
of ILLINOIS
PRESS

TABLE OF CONTENTS

PREFACE

It is frequently asserted that those engaged in constructing and using educational tests have not examined the assumptions upon which these instruments are based. In fact, some critics have maintained that research workers in Education were not aware of the assumptions implied in the instruments and procedures which they are accustomed to employ. In his study of "Effect of Practice on Intelligence Tests," Doctor Glick has rendered a valuable service by subjecting assumptions to experimental investigation. Although critical readers may point out certain limitations of the data, the study is convincing. It is obvious that our use of intelligence tests has implied an assumption which is false, and that in consequence many of the scores yielded by these tests have been given an erroneous meaning.

The publication of this account of Doctor Glick's investigation should serve to call attention to the need for explicit recognition and study of the assumptions implied in educational tests. Until this has been done, our use of these instruments is likely to lead to erroneous conclusions.

Walter S. Monroe, *Director.*

April 28, 1925.

EFFECT OF PRACTICE ON INTELLIGENCE TESTS[1]

CHAPTER I

INTRODUCTION

Intelligence tests do not represent the first attempt to measure native ability. Palmistry, phrenology, physiognomy, graphology and many physical tests were attempts at the same thing. Each was greeted with great enthusiasm and was hailed as a means for securing valuable knowledge relative to native ability, until its real worth and validity were determined by experimental methods. When the assumptions of these so-called sciences were experimentally analyzed, they were removed from the realm of practical science and relegated to the domain of the quack. Group intelligence tests have recently attained great popularity, but we are just beginning to examine critically the assumptions upon which they are based.

Assumptions underlying intelligence tests. Because of the fact that intelligence tests measure native capacity only in terms of behavior, it follows that such measurement must be indirect. All indirect measurements involve assumptions that need to be examined carefully. Among the assumptions implied in our present procedures for the measurement of intelligence are the following:

I. It is assumed that all persons tested have had practically identical environment and equal opportunity to acquire the abilities for which a test calls.

II. It is assumed that the physical, mental, and emotional status of the different subjects is practically uniform and constant.

III. It is assumed that initiative, determination, perseverance, and other similar qualities which are usually considered essential to success, but which it is not claimed our tests meas-

[1]This report has been prepared with "liberal editing" from a manuscript submitted by Dr. H. N. Glick in partial fulfillment of the requirements for the degree of doctor of philosophy in Education in the Graduate School of the University of Illinois, 1924. A number of tables and discussions of minor phases of the study have been omitted. A copy of the original report is on file in the University of Illinois Library. WALTER S. MONROE, *Director*, Bureau of Education Research, University of Illinois.

ure, either approximate a perfect correlation with the traits measured, or do not affect the performances which the tests require.

IV. It is assumed that the functioning of the abilities for which a test calls can be secured at any time and that they are not influenced by the functioning of other abilities.

V. It is assumed that general testing conditions can be controlled.

VI. It is assumed that an intelligence test score is not materially increased by practice or coaching.

Purpose of this investigation. The purpose of this study is to investigate the validity of the last of the assumptions listed; that is, *an intelligence test score is not materially increased by practice or coaching.*

General procedure employed. A procedure was devised for securing a measure of the effect of practice upon (1) the accuracy[2] of the pupil's performance, and (2) the rate of his performance. Two types of practice were used: (1) repetition of exercises similar to, but not identical with, those of the test used (practice without coaching), and (2) deliberate coaching for the tests (practice with coaching).

Varying effect of practice. Investigations of the effect of practice show that the amount of improvement varies greatly. For example, in the case of pitch discrimination, practice produces comparatively little improvement. On the other hand, improvement of more than 1000 percent has been shown in the case of mirror drawing. It appears therefore that we have no general basis for predicting the amount of practice effect in a particular case, and that in order to ascertain such an amount it is necessary to institute a special inquiry.

Practice with identical material versus practice with similar material. We have practice with identical material in learning to operate a typewriter or a telegraph instrument or in learning to play a musical instrument. In such learning the object is to acquire skill in the performance of certain specified exercises.

Practice with similar material occurs in such subjects as arithmetic, algebra, and foreign languages. As the result of practice, a

[2]In this report the term "accuracy" has a somewhat restricted meaning. The "accuracy" of the pupil's performance is measured by the number of exercises which he does correctly.

student is expected to acquire skill in doing exercises similar to, but not identical with, those done during the period of practice.

Since it is the purpose of this study to ascertain the effect of practice resulting from the taking of intelligence tests, similar material was used. The use of identical material for practice would have been unfair to our present intelligence tests, because it is assumed that the subjects tested have no previous knowledge of the particular exercises which they are asked to do. In fact, in most cases it is assumed that they have no definite knowledge of the particular kinds of exercises of which the intelligence test is composed.

Initial assumptions. The writer accepts as valid two conclusions of biology and psychology: (1) that general intelligence or native ability exists, and (2) that general intelligence varies with individuals. He also accepts, with certain reservations, the assumption that intelligence tests measure native ability.

CHAPTER II

EXPERIMENTAL PROCEDURE

Subjects used. The subjects used in this investigation were as follows: forty-five students in the seventh and eighth grades of the Thornburn School, Urbana, Illinois; eighty-five high-school students, Urbana, Illinois; and thirty-five college students of the Massachusetts Agricultural College, Amherst, Massachusetts.[1] Twenty-seven of these subjects did not complete all of the tests and their scores are not included in this report.

Tests used to measure intelligence. Forms 5, 6, 7, 8, and 9 of the Army Alpha Intelligence Examination were used to measure the intelligence of the subjects.

Practice materials. The writer prepared exercises for practice which were similar to, but not identical with, those of the sub-tests of the Army Alpha Intelligence Examination. It was intended to have the practice exercises equivalent in difficulty to the corresponding Alpha tests but there is no experimental proof that these intentions were realized. Twenty practice forms were prepared but only fifteen were administered because, by the time this number had been used, it appeared that the practice had been carried sufficiently far for the purpose of this investigation.

In constructing the practice forms an effort was made to exclude all exercises that appeared in any of the Alpha forms. In a few instances the same exercises were used in two or more of the practice forms. The number of items in each sub-test of the practice forms was the same as in the corresponding sub-test of the Alpha forms, with the exception of Sub-test 3, in which fourteen exercises were used instead of sixteen. This change was made because no more than fourteen exercises could be conveniently mimeographed on one page.

The administration of the experiment. The writer administered all of the Alpha forms, as well as the practice forms. The collection of data extended from October 9, 1922, to May 11, 1923. The sub-

[1]The writer acknowledges his indebtedness to Superintendent William Harris, Urbana Public Schools, Principal M. L. Flaningam, Urbana High School, and Principal R. A. Garrett, Thornburn School, for their assistance and cooperation. The students of the Massachusetts Agricultural College were members of the writer's class.

jects were handled in groups ranging in size from twelve to twenty-five. In the following tables some of the groups include more than twenty-five subjects. In such cases the subjects were divided into two sections for the administration of the tests and practice exercises, and an effort was made to keep all testing conditions constant, except the time of day which in no case varied more than two hours.

The general plan of the experiment was to begin by administering one of the Alpha forms. This was followed on successive days by the administration of the practice forms with the other Alpha forms being given at more or less regular intervals. It was decided more or less arbitrarily that the interval between the administration of the several forms should be one day, with the exception of Saturday, Sunday, and holidays. The work was interrupted by only two holidays and these interruptions affected only two groups. The order of the Alpha forms was varied to correct for any differences in difficulty.

Before the administration of the first Alpha form, the subjects were given but little exact information concerning the nature and purpose of the work. It was feared that some might not make a diligent effort on the first trial if they knew that the purpose of the work was to determine the amount of practice effect. After the administration of the first Alpha form, the purpose of the investigation was carefully explained and all students were urged to improve their scores as much as possible.

The instructions for each Alpha sub-test were given in full on the first trial; but, except for the first sub-test, were omitted on subsequent trials. Very brief instructions were given for the first practice forms. The omission of instructions doubtless put the subjects to some disadvantage but the effect will be to increase the validity of the findings. In the practice "without coaching," the subjects were given no explanation of the method of scoring or of the general principles involved in the tests. In the "practice with coaching," the principles of the test were explained and shortcuts for doing exercises were pointed out. All questions raised by the students were answered.

Attitude of subjects. The attitude of the subjects toward the tests varied. Some were very cautious and did carefully all that they attempted. Others were inclined to sacrifice accuracy for rate of work and evidently resorted to guessing at times, especially when a guess would stand a chance of being correct.

It was anticipated that subjects would grow exceedingly weary of the work before the end of the four weeks of daily testing, and in order to offset this tendency a variety of incentives was introduced. The subjects were told of their scores on the Alpha forms and were encouraged to attempt to increase their scores at the next trial. Treats in the form of candy were frequently distributed, both in the Thornburn School and in the high school. In addition, the subjects in the Thornburn School were promised fifty cents if they continued the work to the end of the fourth week. No tangible incentive was offered to the college students, but all were members of the writer's classes in education and appeared to be interested in improving their scores. Under these conditions an expression of weariness of the task was very unusual. In fact a number of the subjects expressed regret when the work was completed.

Method of measuring rate of performance. One of the fundamental requirements of test construction is that "the test should provide adequate opportunity for all pupils to demonstrate their abilities in the field defined by its function."[2] It follows that the time limit for a rate test should be such that very few, if any, of the subjects will do all of the exercises. Seven of the eight sub-tests of the Army Alpha Intelligence Examination are rate tests, and after practice, only one subject failed to finish some of the sub-tests in less than the time allowed, two subjects finished the sub-tests in less than half the time allowed, and a number finished in slightly more than half time. It therefore was necessary to devise some means for securing a record of the time actually used by a subject when he completed the sub-test in less than the standard time allowed. To accomplish this, a large clock was always started at zero time for each sub-test, and the subjects were instructed that, if they should finish any test before time was called, they should read the clock to the nearest second and record the time at the bottom of the test.

This method of having each subject record his own time may be questioned, because it involves opportunity for dishonesty. In order to reduce the amount of cheating to a minimum, the records of the subjects were checked by the examiner, who, when he saw a subject look at the clock and record the time, would also record the time after the subject's name. Although a record for each subject

---

[2]Monroe, Walter S. The Theory of Educational Measurements. New York: Houghton Mifflin Company, 1923. p. 65.

was not obtained each day, sufficient samples were secured to postulate with considerable certainty the accuracy of the records made by the subjects. Only three instances were found where the record of the examiner did not tally within two seconds that of the subject.

Statistical treatment of data. The score yielded by the regular method of scoring is called the "accuracy score." The total time consumed in completing the several Alpha sub-tests is called the "rate score." A subject's rate score and accuracy score is combined into a single measure, the "corrected score."[3]

The forms of the Army Alpha Intelligence Examination which were used are known to yield scores that are somewhat lacking in equivalence. However, investigation revealed that this lack of equivalence resulted in errors which could be safely neglected in the comparisons made in this study.

The fact, that on the first trial the subjects in general did not attempt all of the exercises of a sub-test in the time allowed and that after practice they generally completed a test in less than the regular time allowance, made it difficult to compute the percent of increase in the rate score. For example, Subject No. 7, Group I, attempted fifteen of the twenty problems of the second Alpha subtest and did nine correctly. On the last trial she completed all of the twenty problems in three minutes and eight seconds and did all of them correctly. Obviously these two records are not directly comparable. It is necessary that both be expressed in terms of either the number of examples attempted or the time consumed. Two procedures for securing an initial rate score were considered: first, to compute the probable time that would have been required to complete the sub-test on the first trial; second, to use the standard time allowance as the initial rate score.

The first procedure is open to the objection that most of the sub-tests are scaled. For this reason it is likely that the pupil's actual rate of work throughout the test tends to decrease as he advances to the more difficult exercises. It would therefore have been very difficult to estimate at all accurately the probable time required for a

[3]The "corrected score" was derived by weighting the accuracy score in proportion to the time not consumed. For example, if a score of 10 was made in two minutes when the standard time allowed was four minutes, the "corrected score" would be 20. This method is based upon the assumption that, if a sufficient number of exercises of the same difficulty had been supplied, the subject would have maintained the same rate of performance for the total time that he did for the actual time consumed.

subject to complete a sub-test on the first trial. Disregarding the scaled structure of the sub-test would result in introducing a positive error in the amount of practice effect.

The second method implies the assumption that the subject did all of the exercises of a sub-test on the first trial. This is not true. In fact several of the subjects failed to complete as many as half of the exercises on the first trial. However, the second method introduces a negative error in the amount of practice effect. As we shall show later, the effect of the presence of such an error is to increase the validity of the conclusions reached. For this reason, this method was used in preference to the one described in the preceding paragraph.

CHAPTER III

EFFECT OF PRACTICE

Distribution of testing and practice without coaching. The distribution of testing and practice without coaching is shown in Table I. It should be read as follows: Group I consisted of high-school students: five freshmen, five sophomores, and two juniors. (Two subjects failed to complete the experiment and their records are not included.) At the beginning of the experiment, they were given Form 5 of the Army Alpha Intelligence Examination. Following this, eight days were devoted to practice which consisted of administering tests similar to, but not identical with, any of the forms of the Army Alpha Intelligence Examination. Then Form 6 was administered, followed by three days of practice and so on. For this group the experiment really closed with the administration of Form 9. The data for the other groups are to be read in the same way. It will be noted that there was some variation in the length of the periods of practice for the different groups.

Gains due to practice. Table II presents a summary statement of the average gains[1] made by the five groups that received practice without coaching. In computing the number of periods of practice given in the second column of the table, the "trials" between the first and last are included. It should also be noted that the scores made on Form 8 were not used in the case of Groups I, II, and IV. The "accuracy score" has been defined as the score obtained by the regular method. In other words, it is the number of exercises done correctly. Table II is to be read as follows: At the end of the experiment, the average accuracy score of Group I was 35 points greater than at the beginning, (absolute gain). This represents an increase of 30.1 percent over the average initial score, (relative gain). The "absolute gain" in "rate score" is 5:25 (read 5 minutes and 25 seconds) and the relative gain is 27.8 percent. For the "corrected score" the two measures of gain are 111.6 percent and 88.4 percent.

In interpreting the facts given in this table, it should be noted that on the final testing (fourth trial) many of the subjects finished

[1]The median gains were also computed, but since they did not differ materially from the average, they are omitted from the report.

TABLE I. EXPERIMENTAL PROCEDURE FOR GROUPS RECEIVING PRACTICE WITHOUT COACHING

Group	Composition of Group	Net Number of Subjects	Order of Experimentation								
			1st Trial	No. of Practice Forms	2nd Trial	No. of Practice Forms	3rd Trial	No. of Practice Forms	4th Trial	No. of Practice Forms	5th Trial
			Form		Form		Form		Form		
I	H. S. Students Fresh. 5 Soph. 5 Juniors 2	10	5	8	6	3	7	4	9		8*
II	H. S. Students Fresh. 7 Soph. 3 Juniors 2 Seniors 1	13	9	4	7	6	6	4	5		8†
III	H. S. Students Fresh 14 Soph. 3 Juniors 1	18	9	4	7	4	6	2	5	4	8
IV	Thornburn School 7th Grade 12 8th Grade 12	22	9	4	7	4	6	3	5	4	8‡
V	Coll. Students Soph. 1 Juniors 17 Seniors 16 Gr. St. 1	35	7		8	2	9	3	5	4	6

*Form 8 was given 72 days after Form 9, but the scores yielded by Form 8 were not used in measuring the effect of practice.
†Form 8 was given 42 days after Form 5, but the scores yielded by Form 8 were not used in computing the practice effect.
‡The scores yielded by Form 8 were not used in measuring the practice effect because of the prevalence of unusual testing conditions at the time of giving this test and during the four preceding practice periods.

TABLE II. AVERAGE, ABSOLUTE AND RELATIVE GAIN IN ACCURACY, RATE AND CORRECTED SCORES (PRACTICE WITHOUT COACHING)

Group	Number of Days of Practice	Accuracy Score		Rate Score		Corrected Score	
		Absolute Gain	Relative Gain*	Absolute Gain	Relative Gain	Absolute Gain	Relative Gain
I	19	35.0	30.1	5:25	27.8	111.6	88.4
II	19	36.9	44.1	4:21	22.3	84.0	90.4
III	19	36.7	33.5	5:20	27.3	104.8	87.6
IV	20	36.5	42.3	3:41	19.1	74.6	79.4
V	14	28.7	19.3	5:06	26.8	99.0	57.9
Total	16.5	33.7	31.4	4:38	24.6	83.8	75.8

*All relative gains are expressed in terms of percent.

some of the sub-tests in less than the regular time allowance and for this reason the accuracy score does not furnish a true measure of the effect of the practice. A measure of the decrease in the time required for completing the Army Alpha Intelligence Examination is given by the average gain in rate which is 5:25 for Group I. An approximate interpretation of this statement is that on the average the subjects of Group I completed the sub-tests in five minutes and twenty-five seconds less than the regular time allowance. Since on the first trial, few of the subjects completed all of the exercises of the sub-tests within the time allowed, this "gain in rate" does not give us a true measure of the effect of practice upon the rate of work on the test. The "corrected score" gives a more truthful statement of the effect of practice, and as might be expected, the gains for this score are larger than for either the "accuracy score" or the "rate score," though still not large enough.

The "corrected score" does not tell the whole truth, because it does not take into account the fact that on the first trial most of the subjects did not complete the sub-tests within the time allowed. The average gain in rate for the five groups combined was estimated to be 9:58, instead of 4:37, as shown in Table II. The gain for the corresponding corrected score is 162.9 points or a relative gain of 131.7 percent, instead of 93.8 points and 75.8 percent. Obviously the average gains shown in Table II are considerably smaller than the real gains. This limitation of the data, however, is not a serious one because the gains given are relatively large.

It is obvious from the facts given in Table II that practice without coaching results in very material increases in the scores made

on an intelligence test of the type represented by the Army Alpha Intelligence Examination. The average corrected scores for the three groups that had nineteen periods of practice show gains of from 84 to 111 points. With the exception of Group V, which had a relatively large average initial score, the gains are in excess of 75 percent of the initial score. Since the method of computing the effect of practice minimized its magnitude, it appears probable that, if a true measure of the effect of practice had been secured, considerably more than half of the subjects would have been found to have doubled their initial scores as the result of approximately seven hours of practice.

Although no specific attempt was made to investigate the question, some data were secured in the course of the experiment which indicated that the limit of the effect of practice was not reached by the end of the fourth week. Hence, if additional practice had been given the subjects, it is likely that some additional gains would have been made.

The distribution of testing and practice with coaching. It was the original intention to confine this experiment to the determination of the effects of practice without coaching but, in the course of the work, the subjects asked so many questions concerning the nature of the exercises of the sub-tests and the procedure in doing them that it was decided to give two groups practice with coaching. The first of these, which is called Group VI, consisted of thirty-three subjects in the Urbana High School. Twenty-six completed the work: one senior, eight juniors, seven sophomores, and ten freshmen. Group VII consisted of twenty-four subjects in the Thornburn School. Twenty-two completed the work: eleven seventh-grade and eleven eighth-grade pupils.

The same experimental procedure was followed for both groups. Form 7 of the Army Alpha Intelligence Examination was given at the beginning of the experiment. On the second day, a half hour was devoted to an explanation of the method of scoring and a discussion of the principles and "shortcuts" relating to Sub-tests No. 1 (Instructions Test) and No. 5 (True-False). All questions that the subjects cared to ask were answered. On the third day, Form 5 was administered. The fourth day was devoted to coaching on Sub-tests No. 2 (Problems) and No. 6 (Number Composition). The fifth day was devoted to practice with a review of the instructions previously given.

TABLE III. EFFECT OF "PRACTICE WITHOUT COACHING" COMPARED WITH EFFECT OF "PRACTICE WITH COACHING" (PERIOD OF PRACTICE TWO WEEKS)

Groups	Number of Subjects	Accuracy Score		Rate Score		Corrected Score	
		Absolute Gain	Relative Gain*	Absolute Gain	Relative Gain	Absolute Gain	Relative Gain
I and II Without Coaching	27	25.5	21.3	3:58	20.2	72.7	59.2
VI With Coaching	26	36.6	32.2	2:44	14.2	71.7	58.4
IV Without Coaching	22	22.9	25.5	3:35	18.4	53.2	56.1
VII With Coaching	17	33.6	38.2	2:10	10.8	53.8	60.3
I, II, IV Without Coaching	49	24.6	23.1	3:50	9.9	65.17	59.7
VI, VII With Coaching	43	35.4	34.6	2:30	12.8	64.6	59.2

*All relative gains are expressed in terms of percent.

Form 9 was administered on the sixth day and Form 6 on the eighth day. The seventh and ninth days were devoted to coaching and practice on some of the most difficult exercises. Form 8 was given on the tenth day. The periods devoted to practice varied from twenty-five to thirty minutes.

In order to provide data for comparison with the gains made by these two groups, the gains made by three other groups were calculated at the end of the second week of the experiment. In Table III, the gains for Groups I and II have been combined so that comparison may be made with the gains for Group VI. The average initial score of Groups I and II combined was 119.5, and that of Group VI, 122.7. Even this difference tends to become insignificant when the differences in the difficulty of the forms of the Army Alpha Intelligence Examination upon which these gains are based are considered. Hence, we may consider Groups I and II

TABLE IV. AVERAGE GAINS ON THE SEPARATE SUB-TESTS (ALL GROUPS COMBINED)

Sub-test	Accuracy Score		Rate Score		Corrected Score	
	Absolute Gain	Relative Gain*	Absolute Gain	Relative Gain	Absolute Gain	Relative Gain
1	3.61	49.8				
2	3.57	33.1	1:23.4	27.9	9.85	85.4
3	2.32	23.4	7.4	8.3	4.08	40.3
4	2.06	15.9	13.9	15.4	7.23	47.5
5	3.93	30.8	36.5	31.0	14.03	86.2
6	5.67	55.4	31.6	17.6	10.25	99.1
7	8.55	34.6	29.7	16.6	11.87	70.4
8	2.66	12.7	57.8	27.0	17.10	62.4

*All relative gains are expressed in terms of percent.

comparable with Group VI and Group IV comparable with Group VII.

An inspection of Table III reveals the fact that in every instance the groups which received "practice with coaching" made greater gains in accuracy but less in rate than those which received "practice without coaching." This superiority in accuracy exhibited by the groups which received "practice with coaching" is doubtless due to the fact that these subjects had a better understanding of the types of exercises which made up the several sub-tests. Their inferiority in rate was probably due to conscious attempts to apply what they had learned through coaching. The average gains as measured by "corrected scores" are practically the same for the two types of practice. This fact suggests the statement that "practice without coaching" has approximately the same effect upon the scores yielded by intelligence tests as "practice with coaching," but an analytical study of the data indicates that the latter type of training is likely to produce a distinctly greater increase in the scores yielded by our present intelligence tests.

Effect of practice upon the separate sub-tests. Since a subject's score on the Army Alpha Intelligence Examination is the sum of the scores on eight sub-tests, the question concerning the distribution of the effect of practice naturally arises. Table IV gives the total average gains separately for these sub-tests.[2] As none of the three scores furnishes a very accurate measure of the improvement in a

[2]In computing the averages given in Table IV, the data for all seven groups were included.

subject's performance, it is not possible to make comparison between the results for the different sub-tests. It is, however, obvious that practice affected a subject's score on each of the sub-tests.

Relation of effect of practice to amount of schooling. It is apparent from Table II that very large gains were made by all groups of subjects. In order to determine more accurately the relation of the effect of practice to the amount of schooling, the subjects were classified according to school grade. The crudeness of the measures of the effect of practice tends to destroy the significance of small differences between gains made by different groups, but Table II, as well as the similar table[3] obtained by classifying the subjects according to school grade, suggests that for subjects above the sixth grade the effect of practice is not materially affected by the amount of schooling.

Persistency of practice effect. In order to secure a measure of the persistency of practice effect, Form 8 was given to seven subjects of Group I seventy-three days after the close of the experimental period, and to eleven subjects of Group II forty days after the close of the period of practice. The subjects from Group I showed an average loss of 9.4 points in accuracy and 2:13 in time. The subjects from Group II gained a fraction of a point in accuracy and lost 1:44 in time. Examination of the records of these groups during the period of practice reveals that Group I made a decided gain on the fourth trial of the Army Alpha Intelligence Examination, which was given at the end of the experimental period. This probably accounts in part for the relatively large decrease in the scores made on Form 8, which was administered seventy-three days afterwards.

Five college students, who had an average accuracy score of 187 at the close of the practice in May, 1923, were given the test in the following December. Their average score was approximately the same. It appears therefore that the effect of practice tends to persist. Hence, a subject who has once received practice probably will always make relatively high scores upon an intelligence test of similar type.[4]

---

[3]This table is omitted from this published report.

[4]Forty-three of the pupils, who were in the seventh and eighth grades and the high school at the time of this experiment, were given Form 5 of the Army Alpha Intelligence Examination about the end of February, 1925. This test was not administered by Doctor Glick and some of the other testing conditions were not identical with those of his experiment. Several of these subjects took Form 5 at

TABLE V. CORRELATION OF INTELLIGENCE TEST SCORES WITH SCHOOL MARKS

Measures Correlated	College Juniors and Seniors *35 Subjects*	High School Juniors and Seniors *10 Subjects*	High School Freshmen and Sophomores *31 Subjects*	Seventh and Eighth Grades *22 Subjects*
First accuracy score with average semester grade	+.40±.10	+.75±.09	+.52±.09	+.67±.08
Last accuracy score with average grade	+.52±.09	+.90±.04	+.60±.07	+.70±.07
First rate score with average semester grade	+.10±.12	+.01±.22	−.07±.12	+.33±.13
Last rate score with average semester grade	+.17±.12	+.52±.14	+.27±.11	+.40±.12
First corrected score with average semester grade	+.31±.11	+.69±.11	+.49±.09	+.71±.07
Last corrected score with average semester grade	+.34±.10	+.78±.08	+.46±.09	+.75±.06

Effect of practice "without coaching" upon correlation of test scores with school marks. Table V presents certain coefficients of correlation between intelligence test scores and the average of the school marks received by the subjects at the end of the semester, during which the experiment was carried on. If we compare the coefficients of correlation for the scores resulting from a first trial with the corresponding coefficients of correlation for the last trial, we find that with the exception of one case practice served to increase the degree of correlation. Since the scores for the last trial of the intelligence test involve a variable negative error (see page 12), the coefficients of correlation with average semester grades are somewhat smaller than they would be if "true scores" had been used. Hence, it appears that, as subjects become familiar with an intelligence test, we may expect the scores yielded by such tests to correlate more and more closely with school achievements as measured by semester grades.

---

the beginning of Doctor Glick's experiment. The scores of the others were reduced to the basis of Form 5 before calculating the increase of the scores secured in February, 1925, over those made in the autumn of 1922. The results show that the persistency of practice effect over the period of more than two years was very slight. In other words, the differences between the scores made at this last testing and those made on the first testing, (the one at the beginning of the experiment) were only slightly greater than would have been expected from the fact that the pupils concerned were more than two years older at the time when the last test was given.

Note by Walter S. Monroe, *Director,* Bureau of Educational Research, University of Illinois.

CHAPTER IV

PRACTICAL SIGNIFICANCE OF RESULTS

Use of intelligence tests in determining fitness for college. The data presented in Table II demonstrated that practice with similar material results in very significant increases in the scores made on an intelligence test of the type used in this experiment. This conclusion suggests a question which may be stated as follows: If from seven to ten hours of practice causes a majority of subjects to double their scores on intelligence tests, do these instruments have any value for determining the fitness of candidates for college entrance? The types of material used in intelligence tests and even intelligence tests themselves are now the common property of all who desire them. If such tests are used regularly by an institution to determine the fitness of those who seek entrance, it is reasonable to expect that many candidates will deliberately prepare for the tests. It is evident, from the facts presented in Chapter III, that we must expect material increase in scores to result from general acquaintance with the exercises used in intelligence tests and a much greater increase when there is extended practice or deliberate coaching.

The fact that practice results in increased scores does not necessarily invalidate the measures yielded by general intelligence tests as a basis for college entrance. If all subjects had received the same amount of practice, it is likely that the scores obtained would approach comparability and hence possess validity as measures of general intelligence. This condition is not realized in most groups to which an intelligence test is given. Some of the subjects may have had no experience in taking an intelligence test and most of the types of exercises included in the test may be strange to them. Others may have taken this or a similar test one or more times. A few may have received extended training or coaching.

Data gathered in this study indicate that approximately 70 percent of the maximum increase in scores due to practice is attained on the fifth repetition. This suggests that a partial equalization of practice may be secured by repeating the intelligence tests from three to five times, using different forms and recording only the scores made on the final trial. This statement is supported by the fact that

intelligence scores secured after practice show higher correlations with average school marks.

Correction of norms for practice effect. Since norms for intelligence tests are usually based upon initial scores of unpracticed subjects, it is obvious that such norms will lead to an erroneous interpretation of the scores made by subjects who have received practice. In fact norms determined for first-trial scores are not suitable for interpreting scores made on a second trial of the same test. In the ordinary use of general intelligence tests, no attempt is made to ascertain the amount of practice which the various subjects have received, but in many cases it is likely that at least a few of the subjects have taken an intelligence test on some previous occasion. If there are such subjects in the group tested, it is inappropriate to use our present norms as a basis for interpreting their scores.

The problem here is similar to that noted in connection with the use of tests for determining fitness for college. Probably the best solution would be to determine norms for scores made after a certain amount of practice, say on the fifth trial. Then, when using an intelligence test, it would be administered five times and only the scores from the last trial counted.

BULLETIN NO. 28

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

SOURCE STUDIES IN AMERICAN COLONIAL EDUCATION

THE PRIVATE SCHOOL

By

ROBERT FRANCIS SEYBOLT
Associate Professor of the History of Education
University of Illinois

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1925